



DEPARTAMENTO DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

TP2: Traceroute

Primer cuatrimestre de 2019

Teoría de las Comunicaciones

Integrante	LU	Correo electrónico
Goldstein, Brian	27/14	brai.goldstein@gmail.com



Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

Ciudad Universitaria - (Pabellón I/Planta Baja)
Intendente Güiraldes 2160 - C1428EGA
Ciudad Autónoma de Buenos Aires - Rep. Argentina
Tel/Fax: (54 11) 4576-3359
<http://www.fcen.uba.ar>

Índice

1. Introducción	3
1.1. Marco Teórico	3
1.1.1. protocolo ICMP	3
1.1.2. traceroute	3
1.1.3. detección De Outliers	3
2. Métodos y Condiciones de los Experimentos	5
2.1. herramientas, tecnologías y metodologías	5
2.1.1. Scapy e Implementación	5
2.2. Estructura de Implementación	5
2.3. geolocalización por IP	5
2.3.1. detección de enlaces intercontinentales	6
2.4. Rutas analizadas	6
2.5. Experimentos Planteados	7
2.5.1. Experimento Mapa traceroute	7
2.5.2. Experimento RTT entre saltos	8
2.5.3. Experimento values of deviation	9
3. Resultados	10
3.1. Technion - Israel	10
3.1.1. Observaciones	10
3.1.2. RTT entre saltos	11
3.1.3. Observaciones	11
3.1.4. Values of deviation	12
3.1.5. Observaciones	12
3.1.6. Observaciones generales no enmarcadas en los anteriores experimentos	13
3.1.7. Conclusiones parciales	13
3.2. Univ-Antananarivo - Madagascar	14
3.2.1. Observaciones	14
3.2.2. RTT entre saltos	15
3.2.3. Observaciones	15
3.2.4. deviation values	16
3.2.5. Observaciones	16
3.2.6. Observaciones generales no enmarcadas en los anteriores experimentos	17
3.2.7. Conclusiones parciales	17
3.3. UNSW-Australia	18
3.3.1. Observaciones	18
3.3.2. RTT entre saltos	19
3.3.3. Observaciones	19
3.3.4. Deviation Values	21
3.3.5. Observaciones	21
3.3.6. Observaciones generales no enmarcadas en los anteriores experimentos	21
3.3.7. Conclusiones parciales	22
4. Conclusiones	23

1. Introducción

1.1. Marco Teórico

1.1.1. protocolo ICMP

El protocolo ICMP es un protocolo diseñado para enviar mensajes de error y de control en el marco del protocolo IP. Posibles ejemplos en los que se utilizaría son: si se quiere acceder a un puerto que no está habilitado o un paquete enviado cumplió su tiempo de caducidad. Son usualmente generados con fines de control o diagnóstico o surgen como respuesta a errores en operación sobre IP (como exemplificamos antes).

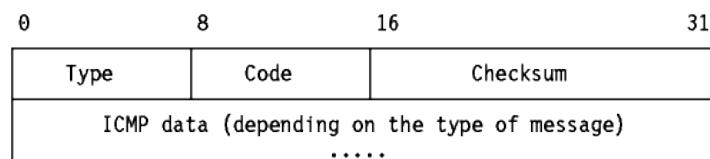


Figura 1: formato paquete ICMP

El formato del paquete es simple, tiene un tipo que especifica qué función de control cumple el paquete, es decir de qué categoría es este paquete, puede ser una respuesta a un echo, puede ser un error disparado por intentar acceder a un destino bloqueado, puede estar informando que caducó un paquete que fue enviado con cierto ttl. El código es una subdivisión de estos tipos. El checksum es el checksum usual y la data dependerá del tipo de paquete si es relevante o no. Para el Tp utilizaremos 3 tipos, el 8 que se interpreta echo request o 'pedido de que te hablen', el 0 echo reply que se interpreta como la respuesta a un echo request, y el 11 que se interpreta como tiempo excedido, este se envía cuando el tiempo de caducidad(ttl) de un paquete llega a 0 y por tanto es necesario avisarle al emisor que su paquete caduca.

1.1.2. traceroute

Traceroute es una herramienta diseñada para analizar las rutas recorridas (serie de IPs) por los paquetes desde un punto hacia otro así como también los tiempos (RTT) entre el origen y cada punto de la ruta.

Hay diferentes formas de implementar traceroute pero la que utilizaremos en el tp es realizar sucesivos echo request con ttl incrementales de modo que el paquete enviado cada paquete enviado llegue un nivel más lejos en la ruta y caduque. Se repite este procedimiento hasta llegar a destino y que te respondan el echo request (o también puede implementarse el final del traceroute leyendo alguna respuesta de ICMP del destino, siempre y cuando no se tenga otra comunicación abierta con el mismo).

1.1.3. detección De Outliers

Outliers son elementos de una muestra tomada de una variable aleatoria que se determina que no es representativo de la muestra. En la práctica suelen ser casos extremos que de algún modo uno puede justificar que pueden desestimarse en un análisis estadístico por lo poco frecuente. Un caso típico de outliers es cuando uno mide una performance de un algoritmo y donde

la performance de miles de ejecuciones es del orden de los milisegundo y uno encuentra una ocasión que tarda minutos, lo razonable en estos casos es suponer descartar este valor por suponer algún problema físico o algo no relacionado al algoritmo.

En este caso buscaremos detectar los outliers de nuestra muestra de tiempos RTT con la esperanza de que estos sean los enlaces intercontinentales. Esperaremos que estos sean identificados como outliers ya que suponemos que sus tiempos (al tener que atravesar largas áreas geográficas) resaltarán del resto de los mucho menores.

Para Detectar Outliers utilizaremos el método de Cimbala:

Este método nos da una receta para decidir si un elemento x_i de una muestra X es o no un Outlier, el método es el siguiente:

1. Calcular la Media ($media(X)$) y el desvío estandar de las muestras(S).
2. calcular para cada elemento x_i su 'absolute value of deviation' $\delta_i = |x_i - media(X)|$, solo analizaremos en esta iteración al elemento con el valor de δ_i mas grande.
3. calcula el valor de 'Modified Thompson tau' (τ), este es un valor calculado en función de la cantidad de elementos que tenemos en cuestión. sera un valor entre 1 y 2 que crece a medida que crece el tamaño de la muestra. La formula para τ es:

$$\tau = \frac{t_{\alpha/2} \cdot (n - 1)}{\sqrt{n} \cdot \sqrt{n - 2 + t_{\alpha/2}^2}}$$

donde n es la cantidad de elementos y $t_{\alpha/2}$ es el valor critico de t de student para $\alpha = 0,05$
4. si $\delta_i > \tau \cdot S$ entonces x_i es Outlier, sino, no lo es.
5. si Resulto que es outlier, retirar este elemento y volver a repetir todo el proceso recalculando media y desvío estándar. Si resulto no serlo entonces ya se puede afirmar la muestra esta libre de outliers.

2. Métodos y Condiciones de los Experimentos

2.1. herramientas, tecnologías y metodologías

2.1.1. Scapy e Implementación

Para la programación del traceroute se utilizo la librería de python scapy. Esta librería brinda una interfaz sencilla para crear recibir y analizar paquetes de las diferentes capas de red. En este TP utilizamos esta librería para crear paquetes ICMP (sobre IP) con ttl (tiempo de caducidad de paquete) incrementales y enviarlos rumbo a diferentes universidades del mundo, de modo que cuando el ttl llegue a 0 el nodo de la red que tenga ese paquete generara un paquete ICMP indicando la caducidad del paquete al emisor del mismo (nuestra computadora).

Luego Una vez recibida la respuesta, también con scapy analizaremos los orígenes de estas respuestas y el RTT (tiempo que tardo en ir el primer paquete + tiempo de vuelta).

Para calcular el RTT por salto se utilizan las propiedades time y sent_time de los paquetes de scapy.

$$rtt = paq_recibido.time - paq_enviado.sent_time$$

Para mayor precisión los mismos paquetes ICMP (con cada valor de ttl) se envían múltiples veces (ráfagas, mínimo 30) y luego a los valores RTT calculados se les aplica la media para calcular los valores medios. Se omitirán los nodos de la ruta que no respondan.

Para calcular los RTT entre saltos, solo es necesario realizar la resta entre RTT de saltos consecutivos. Es importante notar que pueden llegar a darse anomalías de false-RTT y por tanto los RTT entre saltos de negativos. En estos casos optamos por llevar estos valores a 0 y considerarlos despreciables, adicionalmente se excluyen de todo los cálculos de media y desvío estándar para no afectar estas variables artificialmente.

2.2. Estructura de Implementación

La solución pedida se estructuro del siguiente modo:

El ejercicio A y B pedidos por enunciado se implementaron en un archivo python llamado ejercicioAyB.py , al ejecutarse el mismo se pregunta si se quiere efectuar un nuevo traceroute o usar la captura ('./ip_rtts') ya existente. De optarse por realizar un nuevo traceroute se ejecuta el archivo 'tracerouteICMP.py', este genera la captura que luego continuara utilizando el código de 'ejercicioAyB.py' , residualmente tambien se genera un archivo de log para tener control de los paquetes generados en el traceroute.

Una vez ejecutado el 'ejercicioAyB.py' se genera el archivo de 'tiempos_saltos' a partir del cual se pueden ejecutar los experimentos, 'dibujar_mapa.py' para dibujar el mapa y 'graficosRTT.py' para los gráficos de los otros experimentos.

2.3. geolocalizacion por IP

Con el fin de localizar geográficamente las IPs que recorre la ruta se utilizaron diferentes herramientas de geolocalizacion por IP.

La mas importante fue ip2geotools una herramienta que brinda una API que transforma direcciones IP en coordenadas de latitud y longitud. Esta API se utilizo para dibujar los mapas automáticamente a partir de las IPs encontradas en los traceroute.

Como los sistemas de localización por IP no son precisos se utilizaron también complementariamente otros 3 sistemas de geolocalizacion por IP, estos se utilizaron cuando en los experimentos

se querían evaluar posibles fallas de la API ya mencionada. Estas herramientas son IP2Location, ipinfo.io y DB-IP. Todas brindan en mismo servicio de traducción de IP a latitud y longitud, algunas en ocasiones brindan mas información que otras (ciudad, empresa que mantiene el enlace).

muchas veces discrepan y es necesario recurrir a métodos alternativos. Los proveedores del servicio dicen tener una precisión del 95 % para identificar países y una de entre 50 % y 75 % para identificar ciudades (dependiendo si ciudades cercanas se consideran correctas o no).

2.3.1. detección de enlaces intercontinentales

Para la detección analítica de enlaces intercontinentales se plantean 2 métodos, el primero es un método rudimentario que clasificara como outliers a aquellos RTTs que se alejen en mas de 2 desvíos estándar de la media. Planteamos este método para tener poder comparar el método de Cimbala con otro y determinar si es no mas útil (y cuanto mas útil) que un método mas sencillo a la hora de detectar outliers.

El segundo método es el el método de Cimbala casi como esta propuesto en la sección anterior aplicado a los RTTs entre saltos. La única diferencia con el método original de cimbala es que este plantea tomar el valor absoluto de las distancias a las medias para calcular los 'absolute values of deviation' nosotros en contrapartida no calcularemos el valor absoluto ya que no nos interesa encontrar outliers que destaque por ser considerablemente mas chicos que la media. Nos interesan solo encontrar los Outliers que estén por arriba. No tomando valor absoluto nos aseguramos que todo lo que este por debajo de la media quede en el plano negativo y que por lo tanto nunca supere τ (tau).

A la hora de utilizar cualquiera de los 2 métodos no utilizaremos los RTT que nos hayan quedados negativos ya que estos son falsos RTT y no tendría sentido físico tiempos negativos y por otra parte utilizarlos con valor RRT=0 tampoco nos pareció buena idea ya que estaríamos afectando artificialmente los valores de media y desvió estándar.

Notamos que el método rudimentario propuesto es equivalente a designar $\tau = 2$.

2.4. Rutas analizadas

Se analizaron 3 rutas a universidades en diferentes continentes:

- Technion - Israel Institute of Technology , Haifa, Israel , Asia.
www.technion.ac.il
- Université d'Antananarivo, Antananarivo, Madagascar, Africa.
www.univ-antananarivo.mg
- University of New South Wales, Sydney, Australia, Oceania.
www.unsw.edu.au

2.5. Experimentos Planteados

A continuación plantearemos los 3 experimentos realizados (por cada universidad de destino), estos fueron diseñados con la intención de ayudar a interpretar los resultados del traceroute, tanto de manera visual como analítica y poder concluir sobre cuestiones de eficiencia en las rutas, así como también deducir características de la disposición geográfica de los enlaces que se atraviesan.

2.5.1. Experimento Mapa traceroute

En este experimento analizaremos el camino explorado por el traceroute, dibujando en un mapa todas las ubicaciones geográficas que responden un paquete ICMP indicando el time-out de ttl.

Para este experimento se utilizará la API de ip2geotools, un servicio de geolocalización por IP. Este servicio es muy susceptible a fallas por esto, es necesario ser cauteloso a la hora de justificar o rechazar resultados basados en la información que provea, por esto cuando utilicemos la información de este servicio para ayudar a identificar enlaces intercontinentales, reconfirmaremos la información utilizando otros 3 servicios de geolocalización por IP a los que se puede acceder online.

intentaremos observar cuáles son los enlaces intercontinentales así como también sacar conclusiones de que tan eficiente (al menos en sentido de la distancia recorrida) son las rutas que recorren los paquetes, información que podría ser útil en diferentes contextos, como podría ser: optar por replicar servidores en localidades desde las cuales las rutas no son eficientes, dar preferencia a ciertos paquetes (ciertos destinos) en algoritmos de balanceo, etc.



Figura 2: Ejemplo de experimento Mapa de traceroute - vnu: Vietnam

2.5.2. Experimento RTT entre saltos

En este segundo experimento analizando la misma ruta, intentaremos observar el incremento temporal en cada salto de la ruta par luego poder concluir nociones de distancias entre estos saltos y así identificar cuales son probablemente saltos intercontinentales.

Para esto graficamos los saltos de RTT, es decir la diferencia entre los tiempos de RTT entre un salto y su consecutivo (barras celestes). Con el propósito de identificar estos enlaces intercontinentales (outliers desde el punto de vista estadístico), graficamos a modo de referencia la media de la muestra (línea amarilla) y también graficamos la suma de la media +2 desvíos estándar. Esto ultimo, los hacemos proponiendo un método rudimentario para determinar los outliers, serán aquellos que estén a mas de 2 desvíos estándar de la media. En el próximo experimento aprovecharemos la mas sofisticada técnica de Cimbala, técnica que se asemeja mucho a esta técnica rudimentaria de 2 desvíos estándares para muestras muy grandes (el M.T.tau τ se aproxima a 2 para n grande), pero aporta mas precisión para muestras pequeñas (como es el caso de nuestras muestras de saltos de RTT).

Al realizar los experimentos observamos que ciertos RTT entre saltos resultaban negativos, esto se los atribuimos a diferentes motivos, como pueden ser políticas de respuesta a paquetes ICMP, no utilizar la misma ruta de ida que de vuelta, entre otros. Estos valores que dieron RTT negativos fueron asumidos como saltos despreciables y por tanto su RTT se grafica como 0. Para no distorsionar los valores de media y varianza (y por tanto también desvió estándar) con estos RTT negativos (llevados a 0) fueron excluidos de los cálculos de media y desvío. Y se grafican solo a modo informativo pero no aportan información al momento de detección de outliers.

Por ultimo, también graficamos a modo de referencia los tiempos reales de respuesta de cada salto (línea gris). Estos pueden pensarse a priori como un 'acumulado' de los RTT entre saltos, pero formalmente son los tiempos reales de respuesta, la diferencia se hace notar en aquellos saltos con tiempos negativos, ya que el RTT disminuye, algo que no tendría sentido en un 'acumulado'.

A continuación a modo ilustrativo solamente, mostramos un posible gráfico de este experimento:

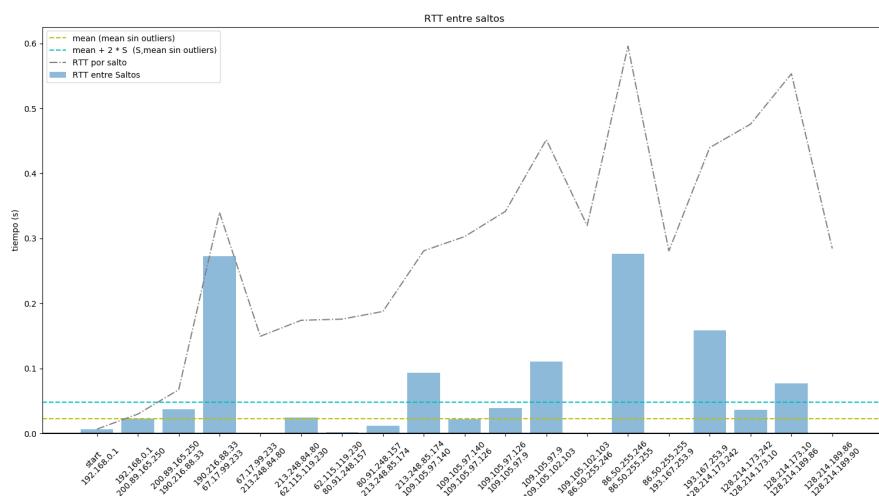


Figura 3: Ejemplo de experimento RTT entre saltos - helsinki: Finlandia

2.5.3. Experimento values of deviation

En este experimento analizaremos los tiempos entre saltos desde una perspectiva estadística para poder detectar los saltos intercontinentales. Con este fin, Utilizaremos el método de Cimbala para detectar Outliers, es decir RTT entre saltos que estén en valores fuera de lo esperado (muy lejos de la media) y como asumimos que estos saltos intercontinentales tendrán saltos de RTT considerablemente mayores que el resto, podremos inferir que estos outliers detectados corresponden a enlaces intercontinentales.

En el siguiente gráfico plasmaremos los 'values of deviation' (distancia de elemento a la media, δ_i) de los tiempos entre saltos dividido la deviacion estándar. Esto lo hacemos ya que según el método de Cimbala un elemento es un outlier cuando su 'value of deviation' supera un cierto valor de $M.T.\tau$ (un numero entre 1 y 2 calculado en función de el tamaño de la muestra, τ) multiplicado por la deviacion estándar ($\delta_i > \tau * S$) o pensado de otro modo (suponiendo varianza no nula, si no no hay outliers), un elemento sera outlier si su $\delta_i/S > \tau$.

Como el proceso de descartar outliers es un proceso iterativo, ya que la media y el desvío estándar va cambiando con cada outlier que se descarta, no serán iguales los valores de δ_i ni de τ ni de S , tal vez hay elementos que en la primera iteración no califican como outliers pero si califican en la ultima, por esto se ilustra también en naranja los valores de δ_i/S de la ultima iteración (aparte de los valores de la primera iteración en azul) ya que estos valores finales, los naranjas, son los que determinaran si el elemento es o no finalmente considerado outlier.

Es importante notar de nuevo como en el experimento anterior que aquellos RTT entre saltos que hayan sido negativos no fueron considerados en el método de detección para no afectar artificialmente el desvío estándar ni la media.

A continuación a modo ilustrativo solamente, mostramos un posible gráfico del experimento:

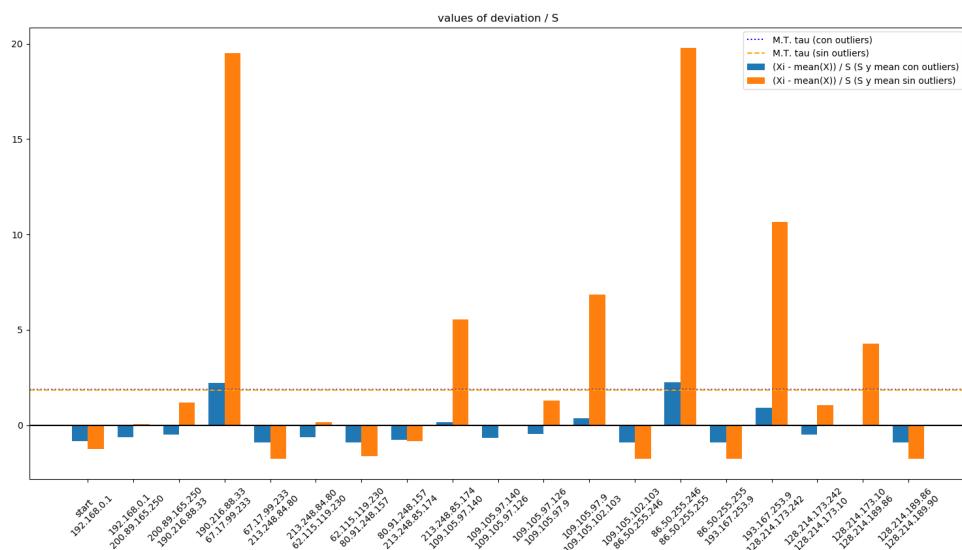


Figura 4: Ejemplo de experimento 'values of Deviation' / S - helsinki: Finlandia

3. Resultados

3.1. Technion - Israel

La primer ruta analizada sera la ruta a la universidad "Technion - Israel institute of technology", esta está ubicada en la ciudad de Haifa, Israel, Asia.

A continuación mostraremos el mapa de la ruta recorrida. Como fue explicado en la sección anterior, el mapa no es 100 % preciso pues para realizarlo se utilizo una herramienta de geolocalizacion por IP con una precisión que si bien es considerable, tiene un margen de error no despreciable.

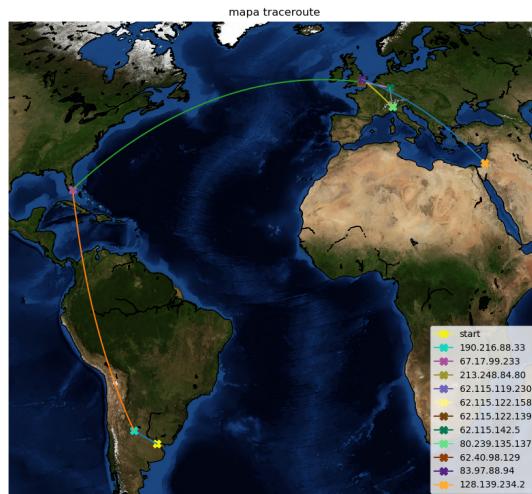


Figura 5: Mapa

3.1.1. Observaciones

Se observa como la ruta seguida es relativamente directa, en el sentido en que, partiendo de argentina (Buenos Aires luego Córdoba), se dirige al norte pasando por los estados unidos para desde ahí cruzar a Europa y luego de algunas redirecciones cercanas (la mayoría locales en reino unido), viaja directo al país de destino.

Se observan También 2 grandes saltos y un tercer salto también considerable. El primer gran salto desde la argentina (córdoba) hasta EE.UU (florida) , el segundo desde los EE.UU. hasta Europa (Reino Unido) y por ultimo el tercer salto considerable, desde allí (UK) hasta Israel.

Se podría anticipar a priori que se cruzaron 3 enlaces intercontinentales. Y los podríamos identificar (nunca perdiendo de vista las consideraciones sobre la precisión remarcadas en la sección anterior) como los enlaces determinados por las siguientes **direcciones IP**:

- Desde: 190.216.88.33 Hasta: 67.17.99.233
- Desde: 67.17.99.233 Hasta: 213.248.84.80
- Desde: 83.97.88.94 Hasta: 128.139.234.2

3.1.2. RTT entre saltos

A continuación se presenta el experimento que expone los RTTs entre saltos del modo en el que fue explicado en la sección 2.5.2. del informe.

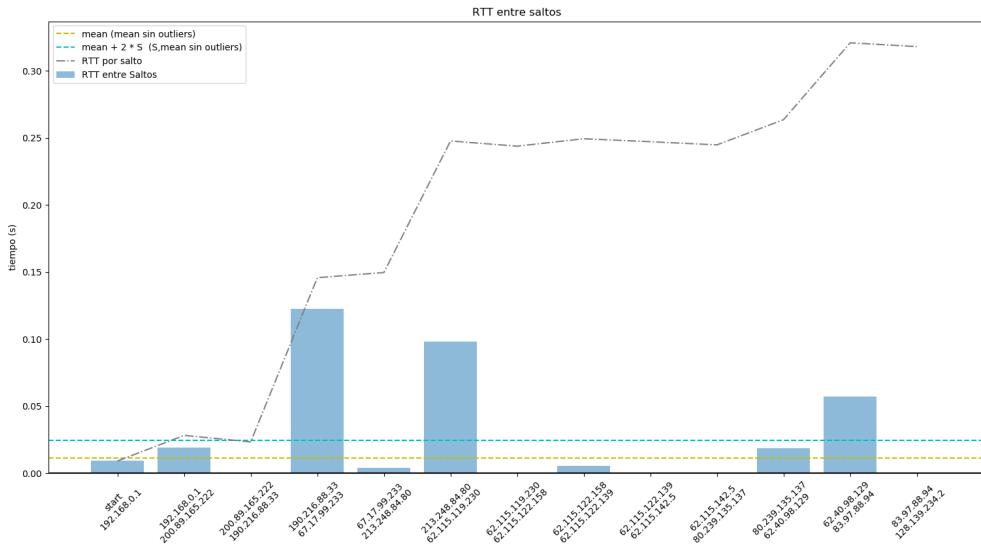


Figura 6: saltos RTT

3.1.3. Observaciones

Se observa que los tiempos reales de respuesta tiene un comportamiento que aproxima muy bien al acumulado de los RTT entre saltos, esto parecería indicar que no hubieron mayores anomalías de aquellas que distorsionan tiempos (denominados False RTT en la bibliografía sugerida [6]).

Se observa en el gráfico que muchos RTT entre saltos tienen tiempos despreciables, se observa como hay 3 tiempos que destacan de los demás, estos 3 superan la media en mas de 2 desvíos estándar por lo que pueden ser considerados (tanto por el método rudimentario propuesto como por el método de Cimbala) como outliers.

De estos 3 saltos destacados solo 1 coincide exactamente con el salto predicho en las observaciones del mapa, el primero.

El segundo destacado se da un salto después de lo predicho, mientras que el tercero se da un salto antes.

Para profundizar el análisis de lo ocurrido (de la discrepancia con el mapa) utilizamos otros 3 sistemas de geolocalización, uno de estos localiza a 213.248.84.80 en EE.UU (en contraposición con Europa, como se predijo anteriormente) explicando así que el salto intercontinental llegue un salto después. Ninguno de los 3 sistemas alternativos pudo justificar que el tercer salto considerable llegue un salto antes de lo predicho, ya que en todos los casos 83.97.88.94 se localiza en Europa (aunque en diferentes lugares). Por lo que solo podemos asumir que el tiempo excesivo o bien se debe a algún factor que nada que tiene que ver con la ruta (tiempos de encolacion, múltiples caminos, etc) o bien también puede ser que efectivamente los sistemas de localización

estén fallando y esta ip este en Israel, creemos esta ultima hipótesis mas viable ya que el tiempo del salto siguiente (el ultimo salto) es despreciable.

3.1.4. Values of deviation

A continuación se expone el gráfico producto del experimento que trata analizar los RTT entre saltos desde la perspectiva estadística, en particular de las variables que influyen en el método de estimación de outliers de Cimbala: los values of deviation, la media y el desvió estándar (en cada iteración del método) con las consideraciones particulares ya desarrolladas en la sección 2.5.3., consideraciones para adaptar el método de Cimbala a la particularidades de este caso de uso (no importan los outliers pequeños, solo los grandes, los valores negativos se ignoran).

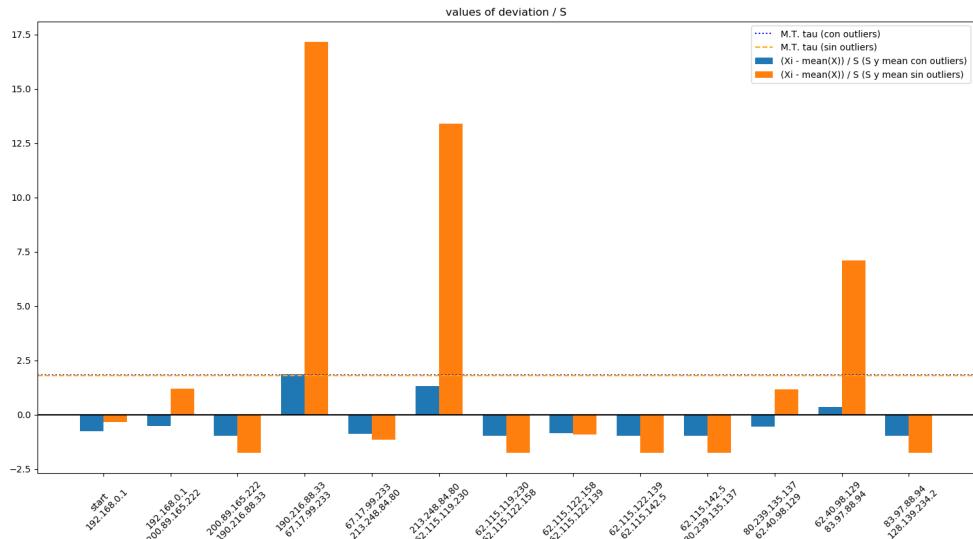


Figura 7: Deviation values

3.1.5. Observaciones

Se observa del gráfico como los valores predichos como outliers en el experimentos anterior con el método rudimentario de 2 desvíos estándar desde la media terminan siendo los mismos en el método Cimbala.

Se observa que si bien en la primer iteración un solo elemento califica como outlier (el salto argentina-EE.UU, en azul) para la ultima iteración los otros 2 saltos importantes son considerados como outliers (en naranja).

Se observa como para aquellos valores de RTT entre saltos por debajo de la media (y por lo tanto negativos en este gráfico) no ganan con las sucesivas iteraciones posibilidad de ser mal clasificados como outliers. Esto es porque a diferencia del método presentado en el documento de Cimbala, aquí no se utiliza el valor absoluto de la distancia a la media, sino que se utiliza la distancia real, esto evita la detección de outliers que se destaqueen por lo poco que tardan, dado

que el problema a resolver era solo detectar aquellos que se destaqueen por tardar mucho para así establecer cuales son los enlaces intercontinentales.

3.1.6. Observaciones generales no enmarcadas en los anteriores experimentos

Se tomo nota del experimento que el ultimo salto que respondió un mensaje ICMP, respondió al mensaje enviado con $ttl = 21$. Teniendo en cuenta que tan solo 13 respondieron, es decir hubo 8 mensajes no respondieron (missing Hops, bibliografía[6]). El porcentaje de respuestas sobre el total fue 62 % (interpretando el total como el largo hasta el ultimo paquete que responde, ya que el missing destination no permite determinar, desde el ultimo respondido en adelante, si no se esta respondiendo o simplemente ya se llego a destino).// La IP a la cual se le enviaron los mensajes (132.68.239.58) no respondió (missing destination, bibliografía[6]).

3.1.7. Conclusiones parciales

Concluimos a parcialmente a partir de lo observado por este experimento que el tanto el método de Cimbala como el método mas rudimentario de clasificar como outlier a aquellas muestras a mas de 2 desvíos estándar de la media, son métodos aparentemente eficientes para identificar enlaces intercontinentales. Por otra parte concluimos que los sistemas de geolocalizacion por IP son útiles para armarse un esquema de la ruta seguida así como también para identificar la cantidad de enlaces intercontinentales y cuales localidades conectan. Concluimos que estos sistemas de geolocalizacion por IP si bien pueden servir para identificar la cantidad de enlaces intercontinentales y sus extremos geográficos, no son una buena herramienta para identificar cuales son las ips en estos extremos, ya que son propensos a fallar en identificar el enlace intercontinental uno (o unos pocos) saltos antes o después de lo real.

De el caso particular de esta ruta, consideramos que hay 3 saltos intercontinentales identificados con las IP que sugiere el estudio de los RTT entre salto (y el método de Cimbala) y estos se dan del modo indicado en el mapa (Argentina - EE.UU, EE.UU-Europa(UK), Europa(UK)-Israel).// Concluimos también que en este caso no parece haber false-RTTs considerables, a pesar de que hubieron algunos valores RTTs que dieron negativos, asumimos por el comportamiento general que eran RTTs despreciables. En cambio si hubo anomalías de missing destination y missing hops.

3.2. Univ-Antananarivo - Madagascar

La segunda ruta analizada sera la ruta a la universidad 'Université d'Antananarivo' , esta está ubicada en la ciudad de Antananarivo, Madagascar.

A continuación mostraremos el mapa de la ruta recorrida. Aplican las mismas consideraciones en cuanto a la precisión de la herramienta de geolocalización por IP que en el análisis de la ruta anterior.

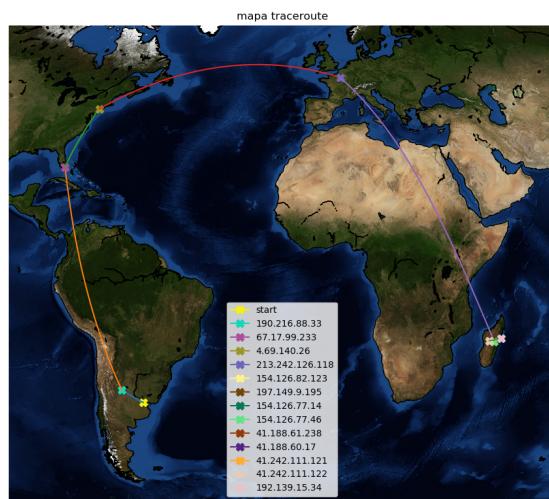


Figura 8: Mapa

3.2.1. Observaciones

Se observa como la ruta sigue una lógica similar al caso anterior, en el sentido en que los mensajes primero se dirigen a EE.UU. para desde ahí cruzar el atlántico hasta Europa de donde es dirigido a destino. Probablemente tiene que ver con la disponibilidad del tendido de enlaces entre estos 2 puntos (EE.UU-Europa) lo que genera el efecto práctico de que las comunicaciones que comuniquen el hemisferio oeste con el este tiendan a en muchos casos comunicarse por este camino (o al menos esto parece cierto para América oriental con Europa, África y Asia occidental).

Se observa un camino relativamente corto en cantidad de saltos hasta llegar al país de destino y en contraposición, una gran cantidad de salto dentro del mismo país. Este comportamiento puede surgir a razón de montones de factores distintos uno de los tantos posibles escenarios puede tener que ver con que el sitio web de esta universidad no tenga mucho tráfico y por tanto no se haya dedicado demasiada atención intentar minimizar el tiempo de acceso al mismo (menos aún desde otros continentes). A propósito de esta última observación y en una nota completamente subjetiva se observa de este sitio web que desde la estética y la funcionalidad aparente, no parece indicar que sea muy utilizado.

Se observan 3 grandes saltos, de los cuales podemos inferir probables enlaces intercontinentales estos son los saltos de Argentina(Córdoba) a EE.UU(Florida), EE.UU(NY) a Francia y de Francia a

Madagascar.

Se observan que las IPs de estos saltos son:

- Desde: 190.216.88.33 Hasta: 67.17.99.233
- Desde: 4.69.140.26 Hasta: 213.242.126.118
- Desde: 213.242.126.118 Hasta: 154.126.82.123

3.2.2. RTT entre saltos

A continuación se presenta el experimento que expone los RTTs entre saltos del modo explicado en la sección 2.5.2. del informe.

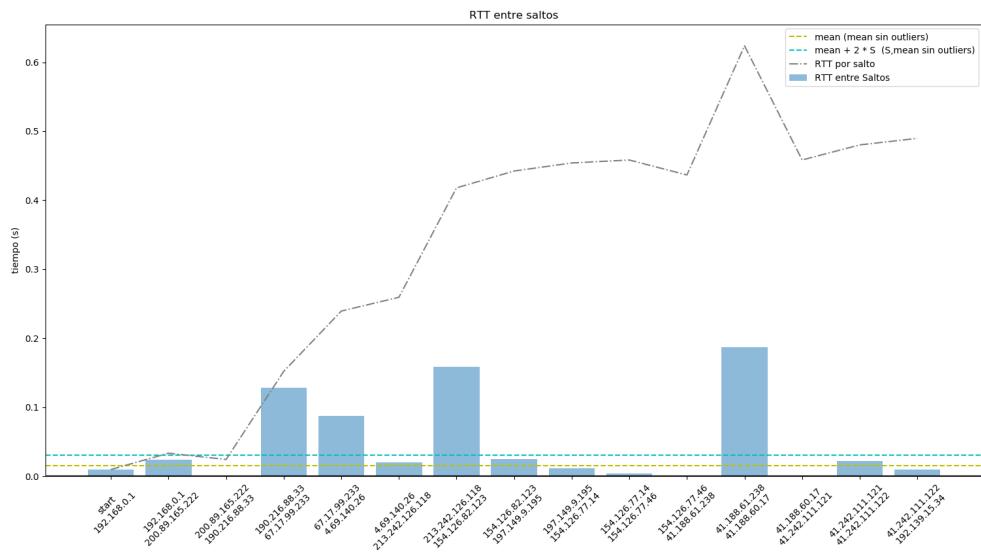


Figura 9: saltos de RTT

3.2.3. Observaciones

Se observa en el gráfico de este experimento como en los tiempos reales de RTT (línea gris) aparecen picos, indicando así caídas significativas en los RTTs entre saltos, esto indica anomalías de la clasificadas false-RTT. En el gráfico de barras estos valores fueron recortados a 0 pues valores de RTT entre saltos no pueden ser validos y también fueron descartados para los cálculos de media y desvió estándar.

Se observa como con el método rudimentario de clasificar como outlier a aquellas muestras que se alejen de la media en 2 desvíos estándar, aparecen 4 outliers.

- El primero de estos coincide exactamente con el primer enlace intercontinental inferido del mapa en el experimento anterior.
- El segundo outlier se da un salto antes de lo inferido en el mapa, para explorar en profundidad esto utilizamos otros 3 servicios de geolocalizacion por IP, uno de estos localiza la

dirección 4.69.140.26 en Francia (en lugar de EE.UU como indica el mapa), adelantando un salto el cruce intercontinental y coincidiendo con lo que indica el método rudimentario de detección de outliers.

- El tercer outlier coincide con el tercer outlier predicho en el experimento del mapa y es el que viaja de Francia a Madagascar.
- El ultimo outlier se da entre 2 saltos localizados ambos en Madagascar, es importante observar que el RTT real cae después de este, este outlier puede deberse a que este salto tarda traspasa (forward) mensaje de manera rápida pero responde lento, aunque podría haber muchas otras explicaciones validas. En cualquier caso consideramos que este es un falso positivo.

3.2.4. deviation values

A continuación se exponen los resultados del experimento que analiza los RTT entre saltos desde la perspectiva estadística, en particular de las variables que influyen en el método de estimación de outliers de Cimbala: los values of deviation, la media y el desvío estándar (en cada iteración del método) con las consideraciones particulares ya desarrolladas en la sección 2.5.3.

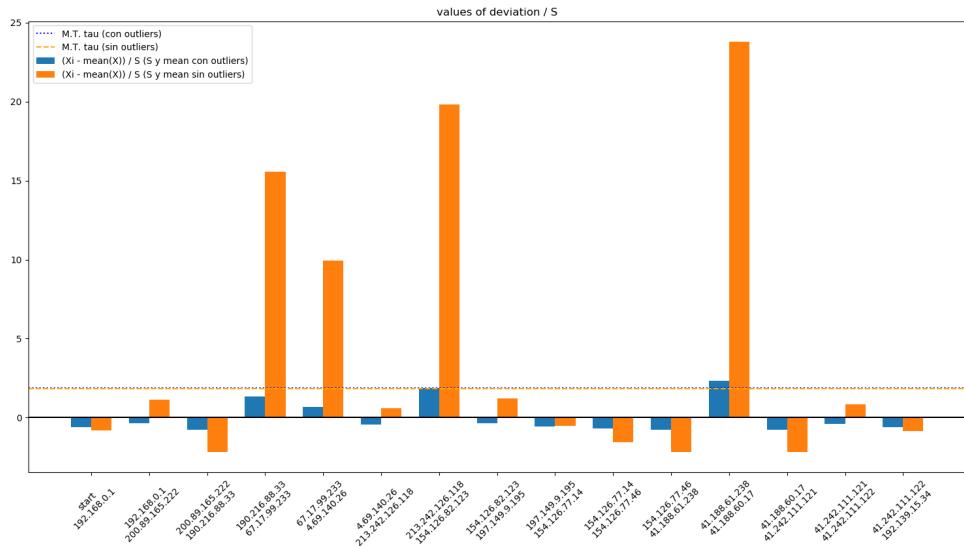


Figura 10: deviation values / standard dev

3.2.5. Observaciones

Se observa en este gráfico como los outliers son los mismos que los estimados con el método rudimentario de clasificar como outliers a aquellos elementos que estén a mas de 2 desavíos de la media.

Se observa como los primeros outliers no son considerados tales por el método en la primera iteración sino que lo son recién en la medida que avanzan las iteraciones y por lo tanto baja la

media y también baja el desvío estándar. Esto ultimo se ve observando que las barras de los primeros 2 outliers no cruzan la linea de τ (tau) pero si la cruzan sus correspondientes barras naranjas.

se observan, que pocos saltos tienen RTTs correspondientes por arriba de la media, lo que nos da alguna pauta de la distribución de los RTTs de una ruta.

3.2.6. Observaciones generales no enmarcadas en los anteriores experimentos

Se tomo nota del experimento que el servidor de destino respondió el mensaje ICMP de time-out, este fue respuesta a un mensaje con ttl = 20, es decir el salto numero 20. Por otra parte solo llegaron respuestas de 15 saltos (5 missing hops) por lo que el porcentaje de saltos que respondieron fue un 75 %.

3.2.7. Conclusiones parciales

A partir de lo experimentado con esta ruta concluimos, que tanto el método Cimbala como el rudimentario (2 desviaciones estándar de la media) son efectivos a la hora de detectar enlaces intercontinentales, sin embargo pueden aparecer falsos positivos cuando por motivos ajenos a las distancias los paquetes tardan de mas en responder. También reafirmamos la conclusión de la experimentación con la ruta anterior en cuanto al tema de los sistemas de geolocalización y afirmamos que son útiles para armar un panorama general de la ruta así como para predecir la cantidad de enlaces intercontinentales, pero no son buenos si lo que se necesita es identificar las ip en los extremos de estos enlaces ya que suelen equivocarse y los enlaces intercontinentales aparecen unos saltos antes o después de los inferido desde el mapa.

Del caso particular de esta concluimos que hay 3 enlaces intercontinentales identificados con las IP especificadas en el experimento de RRT entre saltos. Estos corresponden a los saltos de Argentina a EE.UU., de EE.UU. a Francia y de Francia a Madagascar.

3.3. UNSW-Australia

La segunda ruta analizada sera la ruta a la universidad 'University of New South Wales' , esta está ubicada en la ciudad de Sydney, Australia.

A continuación mostraremos el mapa de la ruta recorrida. Aplican las mismas consideraciones en cuanto a la precisión de la herramienta de geolocalización por IP que en los análisis de las rutas anterior.

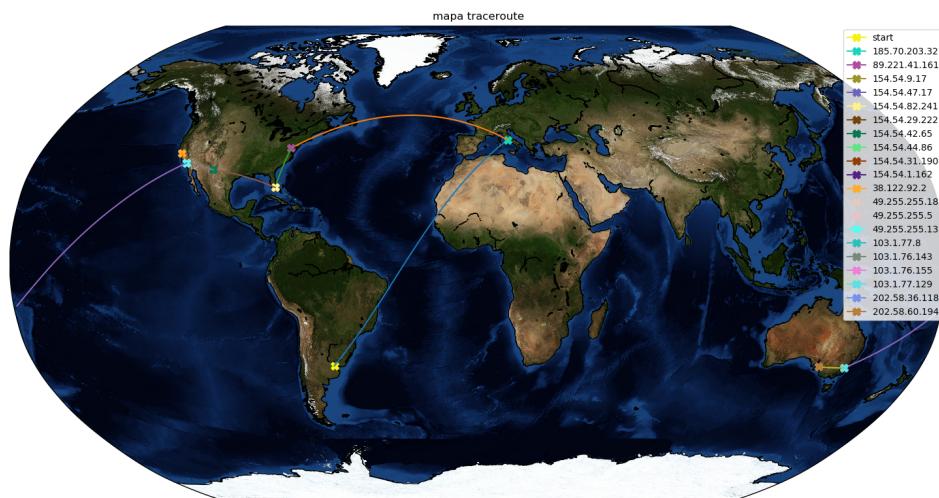


Figura 11: Mapa

3.3.1. Observaciones

Se observa una ruta considerablemente distinta a las anteriores. Mientras que las anteriores desde argentina se dirigían primero a EE.UU para luego cruzar el atlántico y desde Europa llegar a destino, en este caso desde argentina el paquete viaja a Europa de modo directo para luego volver al continente americano vía NY, de allí a florida para luego partir desde la costa oeste hacia Australia a través del Pacifico.

Dado que parece sospechoso el primer salto, es decir, el salto de Argentina a Italia, para luego 'volver' (en algún sentido) a América cotejamos los datos con otros 3 servicios de geolocalizacion por IP, los 3 localizaron la IP 185.70.203.32 (el primer salto) en Italia. Dependemos del posterior análisis de los tiempos entre saltos para esclarecer si efectivamente se esta utilizando un salto intercontinental hacia Europa.

Se observa como siguiendo el patrón de las anteriores rutas, pasa por América del norte a pesar que geográficamente la argentina este mas cerca, seguramente por la disposición del tendido de cables intercontinentales.

Se observan 3 grandes saltos intercontinentales: De Argentina a Italia, De Italia a EE.UU y de EE.UU. hacia Australia. Adicionalmente se observa el camino desde la costa este Americana a la Costa Oeste, este trayecto si bien no es intercontinental podría ser destacado por ser relativamente largo.

Las IPs que suponemos a partir del mapa son enlaces intercontinentales son:

- Desde: 200.89.165.222(start) Hasta: 185.70.203.32
 - Desde: 185.70.203.32 Hasta: 89.221.41.161
 - Desde: 49.255.255.13 Hasta: 103.1.77.8

3.3.2. RTT entre saltos

A continuación se presenta el experimento que expone los RTTs entre saltos del modo explicado en la sección 2.5.2. del informe.

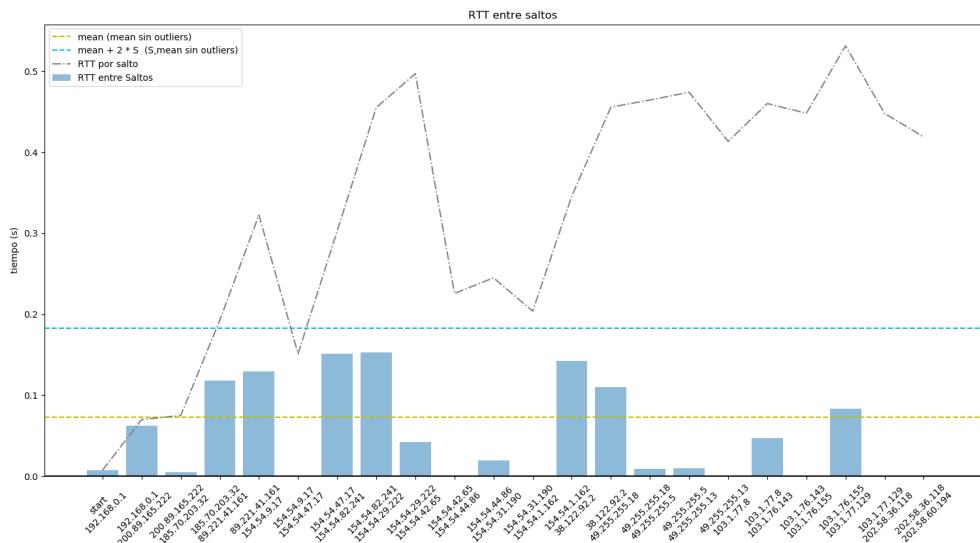


Figura 12: Saltos RTT

3.3.3. Observaciones

Se observa en el gráfico de RTTs por salto (línea gris) muchos picos. Estos picos significan caídas (muy considerables en este caso) en los tiempos acumulados de RTT o lo que es lo mismo la presencia de RTT entre saltos con tiempos negativos. Estos saltos negativos aparecen como consecuencia de anomalías en el traceroute. En particular las anomalías denominadas en la bibliografía [6] como false-RTT se explican de diversas formas aunque una explicación razonable en este caso es que los paquetes no realizan el mismo camino de ida que de vuelta, es decir que si bien los paquetes viajan por la ruta planteada en el mapa, las respuestas por ejemplo desde Estados Unidos, no realizan todo el camino inverso pasando por Europa sino que directamente vuelven a la Argentina por un camino alternativo. Para el análisis, estos false-RTT fueron filtrados y puestos en 0 considerados despreciables y excluidos de los mecanismos de detección de outliers pero al ser tantos y tan considerables no puede asegurarse tan libremente que de verdad los tiempos de los saltos son despreciables y por tanto no puede afirmarse que los resultados serán representativos de la realidad.

Se observa del gráfico barras que efectivamente se destacan de las demás superando la media, sin embargo por la alta varianza de los RTTs ninguna cruza la linea celeste, es decir ningún

RTT esta a mas de 2 desavíos estándar de la media. En este caso el método de detección de outliers rudimentario no detecta ninguno. Como sabemos que efectivamente hay al menos 1 salto intercontinental podemos afirmar que en este caso el método no fue efectivo.

Si bien el método no detecta ningún outlier como salto intercontinental haremos el trabajo de interpretar aquellos saltos con $RTT > 0,1$ ya que consideramos (por lo observado en los experimentos anteriores) que estos son posiblemente los enlaces intercontinentales.

El primer $RTT > 0,1$ de 185.70.203.32 a 89.221.41.161 no coincide con el primer salto predicho en el experimento del mapa, llega un salto tarde. Al segundo RTT considerable de 89.221.41.161 a 154.54.9.17 le sucede lo mismo. Utilizando herramientas de geolocalización alternativas, descubrimos que hay herramientas que ubican al 89.221.41.161 en Italia (en lugar de EE.UU), lamentablemente ninguna herramienta localiza al 185.70.203.32 en Argentina, sin embargo si nos permitimos asumir un error de los sistemas de localización en este aspecto quedarían explicados los primeros 2 atrasos de saltos y se cotejarían como enlaces intercontinentales de Argentina a Italia y de Italia a EE.UU.

A Partir del tercer RTT considerable se hace mas difícil hacer deducciones pues entran en juego las anomalías (como indica el diagrama de RTT 'acumulado' que cae abruptamente) de todos modos intentaremos interpretar los resultados.

El tercer salto considerable parece indicar solamente una recuperación del RTT después de la caída, en este sentido no es un outlier sino mas bien parece ser un síntoma de la anomalía, justificamos esto a partir de que los sistemas de localización no indican un trayecto largo recorrido. El cuarto salto considerable si parece ser un avance territorial que cruza de la costa este a la oeste de los Estados Unidos.

el quinto salto considerable parece ser de nuevo una recuperación, síntoma de la perdida del RTT 'acumulado'. El sexto RTT considerable si indican los servicios de geolocalización un posible enlace intercontinental. Es decir que el enlace estaría de 38.122.92.2 a 49.255.255.18 cruzando de Estados Unidos hacia Australia.

3.3.4. Deviation Values

A continuación se exponen los resultados del experimento que analiza los RTT entre saltos desde la perspectiva estadística, en particular de las variables que influyen en el método de estimación de outliers de Cimbala: los values of deviation, la media y el desvió estándar (en cada iteración del método) con las consideraciones particulares ya desarrolladas en la sección 2.5.3.

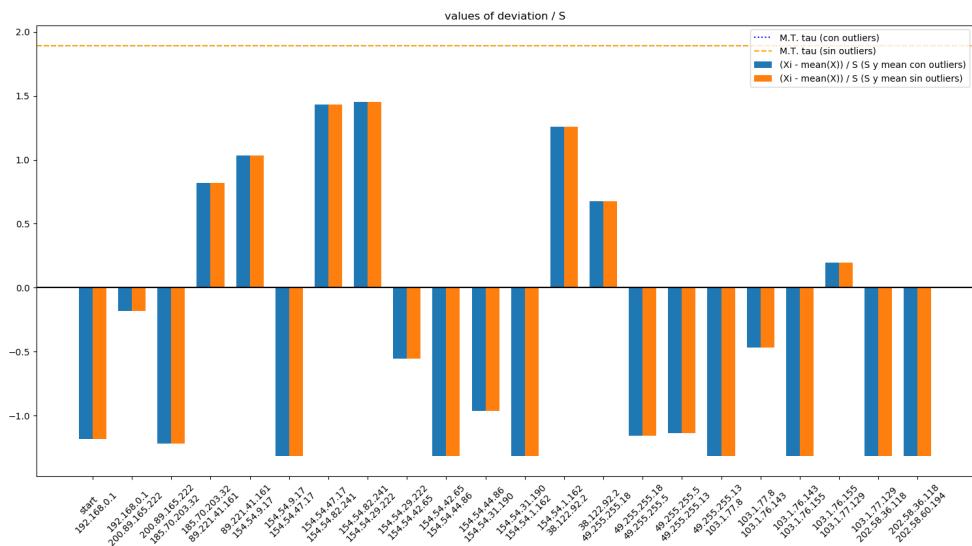


Figura 13: deviations values

3.3.5. Observaciones

Se observa un panorama similar al experimento anterior, el método de Cimbala igual que el método rudimentario de mirar hasta 2 desviaciones estándar de la media aportan los mismos resultados, este resultado es que no se detectan outliers. Como ya analizamos en el experimento pasado los enlaces intercontinentales si están y son identificables pero se requiere un análisis mas profundo. Incluso si la varianza hubiese sido menor y se presentasen outliers el método clasificaría como outliers a los RTTs que provenientes de recuperaciones de perdida de RTT 'acumulado' que se dan por las anomalías de false-RTT del traceroute.

Proponemos como una futura investigación desarrollar algún método que tenga en cuenta los des-balanceos en perdidas de RTT 'acumulado' y posteriores recuperaciones e ignore estas recuperaciones para detectar outliers significativos.

Se observa en este caso que ambas barras, tanto azules como naranjas son iguales, esto es esperable dado que hay una sola iteración de el método Cimbala ya que no se encuentra ningún outlier.

3.3.6. Observaciones generales no enmarcadas en los anteriores experimentos

Se tomo nota del experimento que el servidor de destino respondió el mensaje de time-out esta respuesta fue a un mensaje con ttl=24 , es decir el salto numero 24, por otra parte llegaron

15 respuestas es decir que no respondieron 9, respondió un 62 % (similar a los anteriores).

3.3.7. Conclusiones parciales

Se concluye de lo observado en este experimento que tanto el método rudimentario como el de Cimbala para detección de outliers no es una buen algoritmo para la detección de enlaces intercontinentales ya que no tiene en cuenta (razonablemente pues es una herramienta estadística, no de análisis de rutas) las anomalías que pueden presentarse y como estas pueden a cambio de reducir falsamente algunos RTTs por momentos realzarlos para compensar la perdida de RTT acumulado, por otra parte también puede suceder que ciertas muestras (como fue este caso) presenten medias muy altas y varianzas también muy abiertas, de modo que hayan RTTs que no califiquen como outliers pero si tengan tiempos considerables dado que son enlaces intercontinentales.

De la particularidad de esta ruta si bien ninguno de los métodos analíticos detectó saltos intercontinentales, basta con reducir el grupo de búsqueda a los RTT mayores de 0.1seg y allí aplicar las herramientas de geolocalización para identificarlos como esta explicado en el experimento de RTT entre saltos.

4. Conclusiones

Concluimos a partir de lo experimentado en las diferentes rutas:

- Las locaciones geográficas atravesadas por la ruta en general no están ligadas tanto a la cercanías geográficas como a si a la disposicion de enlaces interoceánicos. Es decir que los paquetes no suelen viajar por el camino mas corto sino que suelen cruzar siempre los mismos países que son aquellos que están mejores conectados internacionalmente.
- Los sistemas de geolocalizacion por IP son una herramienta útil para identificar la cantidad de enlaces intercontinentales , así como también poder plasmar un estimado de la ruta. Sin embargo no son buenos si lo que se quiere es conocer las IP de los extremos de estos enlaces pues suelen equivocarse y adelantarse u atrasarse algunos (pocos) saltos. Recomendamos utilizar alguna herramienta analítica que analice los RTT entre saltos para luego verificar con sistemas de geolocalizacion.
- El método rudimentario de para detección de outliers propuesto (2 desviaciones estándar de la media) parece casi igual de bueno que el de Cimbala, tanto de de un punto de vista teórico así como desde lo observado en los experimentos, la realidad es que por lo menos para los casos probados aquí no hizo diferencia las centésimas que hay de diferencia entre el $\tau(\text{tau})$ y 2.
- Estos métodos de detección de outliers no se adaptan bien a la necesidad de este caso de uso (detección de enlaces intercontinentales) pues al fin y al cabo los tiempos de cruzar el un el enlace suele ser similar cada vez que se cruza. seguramente de mejores resultados establecer un tiempo de RTT fijo a partir del cual se considere como posible enlace intercontinental (algún numero entre 0,05seg y 0,1 seg podría ser un primer punto de partida para buscar el ideal). Otro problema que se puede dar utilizando detección de outliers es que si la ruta es corta y tiene muchos enlaces intercontinentales, no se consideraran outliers porque serian estos los que establezcan en gran medida la media y desvíos estándar.
- Como se concluyo parcialmente luego de estudiar la ruta de la universidad Australiana, detectar RTTs que destaque no es suficiente, es necesario que al algoritmo que se utilice para la detección de enlaces considere factores de anomalías, en particular las anomalías de false-RTT ya que estas corrompen los tiempos tanto reduciendo artificialmente los tiempos como también aveces aumentando los RTTs a modo de compensación por falsos RTT anteriores.
- Aproximamos que de un traceroute son naturales entre 20 % y 30 % de missing Hops. Y no son extraños los missing destination.
- Son frecuentes las anomalías de false-RTT por lo tanto, cualquier algoritmo que planteemos para la detección de enlaces no puede desentenderse de ellos. En contraposición hay evidencia de que sean comunes las anomalías de loops.
- Cuanto mas largas sean las rutas mas propensa a anomalías es la ruta y por tanto mas difícil puede ser identificar enlaces.
- respondiendo a la pregunta de si es preferible darle un numero fijo a τ (tau) , no consideramos que sea demasiado relevante el valor del mismo dado que por la naturaleza del calculo de τ no es un valor que se mueva mucho (suele situarse casi siempre cerca del 1,8)

y la realidad es que los outliers que cruzan ese umbral lo suelen sobrepasar por mucho. y los que no suelen quedarse lejos, no hay gran movimiento de este valor, de hecho por esto se observa en los experimentos que los outliers detectados por este método son los mismos que por el método rudimentario de 2 desviaciones estándares ($\tau = 2$). Si cambiaría como ya concluimos anteriormente utilizar valores de corte fijo para los RTT.